Privacy Schemas and Data Collection: An Ethnographic Account

FINAL REPORT

This research paper reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Eleanor Gerber
Statistical Research Division

USCENSUSBUREAU

Helping You Make Informed Decisions

CONTENTS

EXECUTIVE SUMMARY

This ethnographic research examines a broad range of respondent concerns about privacy and confidentiality. It examines the factors that respondents take into consideration when they are asked to reveal information about themselves across a variety of venues, including censuses and surveys as well as other information gathering forms encountered in daily life. The aim of this research is essentially descriptive. We have portrayed a wide range of beliefs and behaviors which respondents report around the issue of divulging information.

A total of 120 interviews were carried out. Thirty-nine interviews were carried out in Phase I with respondents who had participated in at least one Current Population Survey interview. An additional 81 interviews were carried out in Phase II, with respondents recruited for us by local organizations (many of which had been partnership groups in the census) and by other contacts. Over both Phases, 37 non-Hispanic white, 21 African American, 17 American Indian, 14 Asian, 3 Pacific Islander, 23 Hispanic, and 5 respondents offering more than one race were interviewed.

Semi-structured research protocols were designed to be administered by a team of ethnographers. The interview used flexible probes. The topics included debriefing about Census 2000 and Current Population Survey participation, experiences with other data collections, privacy attitudes, and a series of vignettes. These vignettes served to expand the set of circumstances under discussion to include things of particular interest to the research.

Important findings include the following:

- **Privacy reactions are highly situational.** Respondents decide anew whether to answer questions in each venue where they are encountered. Items that are highly protected in one venue may not be in another.

- **A descriptive model for understanding how respondents decide whether to divulge information was created.** This model includes three main parts: an assessment of the sponsor of the questions; an assessment of whether the questions are relevant to some legitimate purpose of the sponsor; and an assessment of risks and benefits of divulging information.

- **In assessing sponsorship, respondents want to approve of the agency sponsoring the questionnaire.** On the whole, our respondents preferred governmental to commercial sponsors, with the exception of certain government agencies which control negative consequences.

• **Respondents are also aware that sponsorship may be misrepresented; thus, the authenticity of the data collection is an issue for respondents.** This makes telephone mode interviews highly unpopular, because it is impossible to be sure of a caller's true identity.

- **Respondents form expectations of what questions are legitimate for the sponsor to ask, based on their understanding of the nature and purpose of the survey and the sponsoring organization.** Questions that go beyond this framework of expectation may be perceived as intrusive.

- **Respondents evaluate the risks and benefits of providing information.**

- **All respondents are familiar with exchanging information to receive particular benefits, for example, providing information to lenders or to social service agencies in order to receive services.**

- **Respondents are also motivated by altruistic benefits, such as providing information to the decennial census to enable services for a local area.** They also may see participation in surveys and censuses as a way of bringing a group or a point of view greater attention. This is called "having one's voice heard." It was a powerful motivation for Latinos and American Indians to participate in the census.

- **Respondents also worry about the risks of divulging information about themselves and their families.**

- **To respondents, the most important of these consequences are the possibility of fraud (if information they give is used by criminals).** This makes issues of data sharing anxiety provoking. Loss of control of data is worrying in itself.

- **Respondents are also concerned about government agencies which control negative consequences, such as police agencies, the Internal Revenue Service, the Immigration and Naturalization Service, and for American Indian respondents, the Bureau of Indian Affairs.** These agencies are not seen as benign.

- **Because respondents believe that information is freely shared between agencies, despite any assurances of confidentiality, if they have something to hide, they are reluctant to provide it to any government agency.**

This model of how respondents decide to reveal information in censuses and surveys is widely shared in all groups. However some probable differences did emerge. One difference was between more and less technologically sophisticated respondents. Technologically sophisticated

respondents were more comfortable with providing information on the Internet, and felt more able to deal with any potential problems that might occur. Simultaneously, however, such respondents often did not believe that it was possible for any institution to completely assure privacy or confidentiality to persons providing information. Differences in privacy sensitivities also emerged for groups which have had negative experiences with particular agencies of government, such as for legal immigrants who have experienced difficulty in crossing the border.

Recommendations based on this research include:

- **Because privacy judgements are situational, it is not possible to create a list of items that will always or never be considered private.**

- **Because the sense of intrusiveness of questions is situational, be careful how disparate topics are combined in one survey setting (such as topical modules or supplements.)**

- **Include the idea of having one's "voice" heard in motivational material for minority groups.**

- **Through media coverage of other agencies and organizations, respondents are aware that fraud may occur through the action of individual employees. Describe the Census Bureau's internal controls on the handling of data in explanations of confidentiality.**

- **Because respondents' comfort with questions rests on their assessment of the sponsor's legitimate right to know the information requested, provide good, understandable explanations of why these data are needed and how they will be used.**

1. BACKGROUND

Many factors affect the public's response to requests for information in government surveys, including personal experience, cultural value systems, and self-interested or self-protective responses to social circumstances. Expectations are formed through experiences with all data collectors and all modes: school forms, job applications, magazine 'surveys,' phone calls from marketers and the like. Respondents absorb potent images of privacy at risk in fictional accounts and news stories. In undertaking this research, the goal has been to create a preliminary sketch of this wider context, and to locate respondents' reactions to government surveys within it. We found it useful to focus on the decision to provide (or to refuse to provide) information about oneself or one's family; how this decision is constructed, what factors are taken into account, and what other concerns or ideas are evoked in considering this decision. This set of beliefs and connections may be thought of as the schema surrounding privacy. In order to elicit the full web of their ideas, in the most naturalistic way possible, relatively unstructured conversations with respondents were the most appropriate method. [1] Therefore, we have adopted exploratory qualitative techniques for this research. Our aim is not numerical assessment of different points of view, although we were interested in the diversity which arose within our data. Rather, our aim is to portray a spectrum of beliefs and responses to privacy issues, and to show how these concerns are interconnected. [2]

2. METHODS

The research on which this study is based occurred in two phases. In the preliminary phase, we concentrated on privacy issues surrounding demographic surveys with a small number of respondents. The second phase addressed issues of interest to the decennial census.

2.1 Respondents

In Phase I, interviews were carried out with 39 respondents, all of whom had previously participated (at least for one month) in the Current Population Survey (CPS.) (The CPS is an employment survey sponsored by the Bureau of Labor Statistics and the Census Bureau.) Interviews were carried out in Northern Virginia, Los Angeles, the Boston area, and Chicago.

---

[1]Seven anthropologists were involved in this research, including two staff members at the Center for Survey Methods Research and five others who participated under contract. I would like to acknowledge the contributions of Alisu Shoua-Glousberg, Betsy Strick, and Jessica Skolnikff , Susan Trencher, Bhavani Arabandi, and Melinda Crowley to this research.

[2]It should be noted that our aim was not to account for all reasons why respondents refuse surveys or survey questions. Our respondents told us about non-privacy related reasons for refusing to participate in surveys, including time constraints, questionnaire difficulty and the like, but these reasons will not be discussed here.

Twenty seven of the respondents were non-Hispanic White, three were Hispanic, and six were African American and three were Pacific Islander. These interviews took place between June and December of 1999. In Phase II of the research, we focused on the decennial census. We recruited through a wide variety of citizen or social service groups, some of which had served as Partnership Groups in the decennial census. In addition, an ethnographer with ties to a Native American group in Oakland California arranged interviews with us in that community. We also used personal contacts to identify several respondents who could be considered technologically sophisticated. These included individuals working in highly skilled computer jobs (such as consultants, software engineer, etc.) and two individuals who worked for data mining companies. These interviews were carried out between June and October of 2000.

In Phase II, 81 interviews were carried out, in a variety of locations, including Washington DC, Chicago, San Diego, Los Angeles, Oakland, Miami, and Northern Virginia. Fifteen African American, 20 Hispanic, 17 Native American, 14 Asian, Pacific Islander, and 10 non-Hispanic White and 5 multiracial respondents were interviewed.

2.2 Research protocols

For Phase I, a semi-structured research protocol was drawn up for use by the anthropologists connected with this research. Because the aims of the interview were generally exploratory, interviewers used flexible probes to follow up interesting lines of discussion. The Phase I protocol included debriefing about CPS participation, questions about other experiences with requests for information, how respondents decide whether to reveal information, and questions about means of controlling information. Next, a set of eight vignettes was administered, each of which described circumstances in which the central character has to decide whether or not to divulge information. These vignettes served to expand the set of circumstances under discussion to include things of particular interest to the research. Thus, the Internet, revealing information over the telephone, proxying issues, risks associated with giving information, and issues of information sharing and the belief in assurances of confidentiality were suggested by the circumstances of these vignettes. The main aim of these vignettes was to elicit the reasoning processes which respondents applied to the decisions faced by the central character in the vignette (see Gerber, 1994). This protocol was pretested with five interviews.

The Phase II (see Appendix II) research protocol focused on somewhat different issues. We debriefed respondents about their experiences with the Decennial Census, which had taken place only two months prior to the start of ethnographic interviewing. We added questions eliciting respondents' understanding of specific privacy terms and concepts, their sense of whether privacy has increased or decreased, and their reactions to various modes of questionnaire administration.

The vignettes were revised somewhat and new ones were created to elicit responses on several new topics of interest. In particular, we were interested in assessing respondent's knowledge of and reactions to issues of data sharing. [3] This new research protocol was pretested with ten interviews.

3. LIMITS

This research should be taken as a small scale, exploratory study designed to provide insight into the background beliefs and understandings of respondents. Respondent selection in this study, as in most qualitative research, was not part of a representative sample. Therefore, no statistical conclusions should be drawn (and in fact we have not expressed our findings in this way.) The aim has been to portray a range of concerns, rather than to draw valid statistical conclusions about the frequency of these beliefs within the population.

As in all qualitative research, the depth of the interviews precludes collecting data from a large sample of respondents.

One value of qualitative research is that it allows new analyses to emerge which were not thought to be relevant during the planning of the research. This is the case with some of our findings. We did not know that social class or technological awareness would be relevant to our analysis. Therefore, our recruiting plan does not reflect this element, and we may have smaller groups of persons representing the full array of these characteristics than might have been ideal.

An important consideration in assessing these findings is the time frame in which they occurred. The responses to privacy concerns described here were collected prior to important events which have probably affected the way in which respondents think. These include the national responses to the terrorist acts of September 11, 2001 and social and legislative changes which have occurred as a result. Changes in privacy beliefs and behaviors following these events cannot be assessed.

4. FINDINGS

4.1. Respondents' Concept of Privacy

In order to understand the domain of privacy, it is useful to examine the concept in general, from the respondents' viewpoint. This concept, and the specific language used to communicate these ideas, are a natural starting point for this discussion.

[3]The Phase I protocol also used several card sorting tasks. The Phase II protocol included a section assessing respondents' understanding of specific confidentiality language used in demographic surveys and decennial contexts. These data will not be reported here.

4.1.1 Range of privacy vocabulary

In our second research protocol, we asked a series of questions designed to elicit definitions of "privacy" and other concepts which respondents had used in our previous discussions. Some of these concepts, like the idea of what was or was not someone's "business," seemed critically related to the idea of privacy. In addition, the words "private" and "personal" were frequently used, sometimes with contrasting meanings and sometimes as synonyms. We had also noted that the idea of "intrusion" or "invasion" of privacy appeared salient when they were used in the media during Census 2000.

A wide variety of terms were used in discussions of privacy. An extensive, but by no means exhaustive, list of such terms is presented in Appendix I. It is interesting to note that a complex and varied vocabulary exists to express concerns about breaches of privacy, but only a few ways of expressing comfort about information exist. Language exists to express a rough scale of comfort, with terms such as "open," "trusting" and "nothing to hide" on the positive side. Doubts and problems are expressed by terms such as "cautious," "skeptical," "wary," and "leery." The term "paranoid" is also used, either as a self-descriptor ("I'm a little paranoid") or as a way of minimizing one's wariness ("I'm not paranoid, but...")

4.1.2 Basic definitions of p r i v a c y

We were also concerned with respondents' basic definitions of privacy. Our respondents were frequently not able to supply abstract definitions of the concept, and thus this discussion is largely drawn from examples they gave. Appendix I indicates that respondents' sense of privacy is involved with boundaries: either personal ones like "keeping to yourself" or boundaries relating to the household or family, like admonitions not to put private matters "on to the street." Respondents thus are interested in keeping certain information about themselves within a particular range.

In general, the examples and definitions provided by our respondents indicate that the most important dimension of meaning for the term "private" is the effort to control of information. The term "personal" also arose frequently in our interviews. Many respondents are unable to provide an explanation of the difference between this term and "private." In fact there is a large area of overlap between them. However, examples used to illustrate the two indicate some difference. "Personal" implies facts which are closely associated with the self. Examples ranged from items relating to the physical self (one's shampoo or sexual habits) to personal choices (for example, hobbies or interests) to information used as personal identifiers (for example, one's name, race, or social security number). When these bits of information are controlled, they are simultaneously private. Thus, sexual habits and social security numbers are the kind of personal information which is both private and personal. It is important to note, however, that not everything which is considered personal is always highly protected. For example, respondents may see hobbies as personal, but not be particularly invested in protecting that information from revelation. The reverse is not true, however. Things which are highly protected appear to become "personal" even

if they have little initial connection with our physical or mental selves. The best example would be the numbers which are used as unique indentifiers for individuals, such as social security numbers or bank account numbers.

The terms "intrusion" and "intrusive" were of interest to us because of their frequent use in the media, with reference to Census 2000, just prior to our second field period. Most respondents do not use these terms naturally, and the terms had to be introduced. Some respondents associated the terms with press coverage of the census, and referred to questions thought to be unnecessary or "too detailed." Respondents referred to housing questions, income questions, and commuting questions. Since most of our respondents had only seen the short form, they were reporting on what they had read or heard elsewhere. The questions asked on the short form were not considered intrusive. For most other respondents, the term "invasion" or "intrusion" triggered associations with having spatial boundaries broken: people walking into bedrooms unannounced, or peering into houses across the backyard fence, for example.

A much more natural way to describe questions which are regarded as intrusive is to describe them as "none of your business." While almost all respondents found this phrase to be impolite, and denied ever using it in interaction, they recognized the concept as something they might think in response to nosy questions. Information is a questioner's "business" if they can establish a legitimate right to the information. This means that there is a recognized and approved purpose for the questioner to have that information. This assessment forms an important part of the way that respondents decide whether or not to reveal information, to be discussed in the next section.

4.2. A decision model for revealing information in surveys and censuses

This section will describe the general schema which respondents use for deciding which information to reveal about themselves and their families in particular circumstances.

4.2.1 Situational decision making

In planning this research, we began with the naive concept that information on certain identifiable topics would be rarely revealed, and that other topics would be readily revealed in almost all circumstances. For the most part, however, this is not an adequate conceptualization for the way that our respondents dealt with privacy. Instead, they made a complex assessment of who was asking and what the consequences of answering might be, given their own particular circumstances. Thus, information is not private or public in itself, but is revealed or withheld as a result of a situational judgment.

Even simple demographic information can be treated as highly private. For example, an actress told us that she never reveals her age because it may affect her ability to find work. Similarly, even highly sensitive material may be revealed if the situational judgment indicates a need for it. Thus, a number of respondents easily imagined answering survey questions about the number of their sexual partners for a survey with a medical purpose.

4.2.2 Sponsorship and authenticity

In deciding whether to answer, respondents are very concerned with knowing to whom they are giving information. This judgment resolves into two related questions. First, respondents must determine whether or not they approve of the individual or organization collecting the data. We refer to this as the sponsorship of the question. An additional assessment must be made. The questions are often answered through agents which the sponsor has authorized, such as interviewers, or remote collection devices like questions on a website or mailed questionnaires. Thus, for respondents, the authenticity of the agent or collection device presents a second question.

4.2.2.1 Establishing bona fides: authenticity According to some of our respondents, it is impossible to know whether or not someone who asks for information is really who they say they are. This is an example:

> "With what people can do with the computer any more and the way people have found ways to skirt laws concerning impersonating various agencies, its entirely too easy for somebody to put together a form that implies a connection with a legitimate business that isn't really that...and they can name themselves the FBI, which stands for...Fred's Business Institute, and just put FBI at the top."

Being certain of the questioner can become even harder over the telephone, because it is easy to misrepresent an identity there. (This is why some respondents said that they will not answer any telephone survey questions). Some were also aware that an identifying logo on a questionnaire or an ID badge can be faked. Respondents reassured themselves about this in a variety of ways. In the case of government surveys, respondents were looking for something which marked the data collection as "official." Badges were mentioned, and advance letters were viewed as a mark of the serious intent of the sponsor. One respondent said that he looked for some kind of notary seal or watermark to be used on the letter or the survey form itself.

Beyond this, respondents' search for authenticity in sponsorship became more personalized. For some respondents, deciding that an interview was legitimate required an additional personal assessment of the interviewer, based on his/her behavior and bearing. One respondent described this to us as a "leap of trust."

In fact, we were struck with how many of our respondent described their interviewers in very positive terms. Interviewers were described as "nice," "agreeable"and "bubbly." Some respondents indicated that non-substantive, personal interactions were the most memorable thing about their participation in CPS. Respondents recalled jokes that were made, mutual interests in dogs or travel, flattering questions about the home schooling activities of one respondent, and the like. We had the general impression of interviewers attempting to transform an anonymous relationship into a personal one. This may have had the function of preventing boredom and burnout in a set of repetitive questions. It may also have had an effect on respondents' acceptance

of the legitimacy of the interview. Anonymous relations are subject to mistrust, but once they were transformed into personal ones, benefit of the doubt could be given and the "leap of trust" made.

4.2.2.2 Sponsorship Respondents wanted to know who was collecting the information, and whether they approved of the agency in question. If they do not approve, they will not agree to cooperate. For example, many of our respondents did not like marketing research questions, and said they would not answer any questions at all for such a sponsor. One respondent said she would not answer questions for the Centers for Disease Control, because she disliked the research they do on "certain diseases." Respondents did not require specific or accurate information to judge a sponsor, and often used what they were able to deduce from the agency's name. (An example was a respondent who had a positive reaction to a "Health Department" because "health sounds better than disease"). Thus, respondents seem to be forming their judgments of sponsoring organizations, and what they are entitled to ask, on somewhat vague and inferential grounds.

In general, collecting personal information is widely considered a legitimate function, and our respondents concede wide rights to a variety of superordinate authorities to collect it. In fact, we were often struck with the willingness of our respondents to take on the role of such authorities in these agencies when considering whether or not to divulge information. Thus, people reason the insurance company has a right to information about prior health conditions, mortgage lenders should have access to information about your credit history, etc. Respondents even took this attitude with information which they might regard as inappropriate to discuss with acquaintances, if they could see a benign use for the information. One of our vignettes described a preschool which asked parents about how they discipline their children and how often they quarrel. Although this could be easily marked as information that should "stay in the family," many respondents thought that the school might be better able to educate the children if they had this sensitive information.

4.2.2.3 Government sponsors We asked respondents how comfortable they would be in revealing information to particular governmental agencies, including a variety of agencies on a state and local level. In these data, governmental organizations were generally perceived as having helpful or benign goals, and their rights to collect information tended therefore to be accepted. The Census Bureau is widely seen as having benign intentions. Respondents tended to search for good reasons to collect specific data, if they were not immediately apparent, and to conclude "they must have a good reason" to ask. A few government agencies were not granted this credit. These included the Internal Revenue Service and the Immigration and Naturalization Service, police agencies (such as the Federal Bureau of Investigation and local police) and the Bureau of Indian Affairs.

Recruiting in the second phase of research attempted to locate respondents who might have a more negative attitude towards sharing information with government organizations. The mistrust

we encountered varied considerably between groups. Mistrust of government generally was highest among the young, immigrants, African Americans and Native Americans; however , we located these attitudes to some extent in all groups.

This mistrust is generally connected with the consequences that particular agencies control, such as being deported by the INS if one is an undocumented immigrant. However, the mistrust is also connected with a belief that the government is "monitoring" or "tracking" individuals. (Respondents were often familiar with the idea from media sources, but only a minority believed in it. Others thought that the technical potential was there, but were not sure if it really occurred.) The following is an example of a respondent who believes that the Federal Government is tracking individuals. The respondent in this case, is a White middle-class homemaker. Here, the issuance of social security numbers to infants is taken as evidence of government tracking of individuals.

> "The federal government keeps track of everybody...You didn't have to have a social security card until you started work. Now it's required…as soon as a child is born. And you don't think that's a way the government is tracking individuals? You'd better believe they are!"

It is interesting to note that belief in government "monitoring" of individuals doesn't mean automatic refusal to divulge information to government agencies. If respondents believe there is a good purpose to be achieved by giving the information, they'll cooperate despite their suspicions. Thus, in agreeing to cooperate with the census, respondents see the benefits to the community as more salient then the vague and rather distant risk of adding to government files on themselves.

Other ideas support cooperation in the face of significant suspicion about data storage. One in particular was that even if tracking occurs, there would be no reason to single out the respondents' personal data. That is, their data will not call attention to them because their lives are not noteworthy, average or even rather boring. Related to that was the notion that surveillance could turn up nothing that could cause them harm because they "aren't doing anything anyway." This implies a belief that the government is primarily using stored data to find law breakers.

4.2.2.4 Commercial sponsors Although respondents see legitimate sponsors in commercial organizations such as bank and insurance companies, these organizations were not given nearly the same latitude to ask questions as were government agencies. That is, information considered "necessary" to the specific transaction is understood to be the commercial organization's business, even if that information is considered highly sensitive (like income or credit history). Beyond this, respondents are not likely to give the benefit of the doubt, as they are willing to do for agencies like the Census Bureau. For example, many respondents said they answered questions directly related to a product that they have purchased or used, but would not answer ancillary questions, such as those about their lifestyle, preferences, or other purchasing habits. Thus, respondents could see why a company might have a legitimate stake in knowing how satisfied a customer is with a recent purchase, but could not understand why that entitles the company to information about their levels of education.

Respondents resented commercial enterprises that collect information to sell it to others. They often complain about not receiving any profit from information which they regard themselves as owning.[4] In addition, they dislike attempts to collect information to market things to them at a later time, although this tends to be associated with the annoyance they feel at junk mail and junk telephone calls. Many respondents said that they refuse to cooperate with any marketing questions at all.

4.2.3 Relevance of questions to the sponsor's purpose

The decision schema for divulging information required a judgment about the relevance of the specific questions to a legitimate purpose of the questioner. If the requested information was not viewed as relevant to a legitimate purpose, respondents regarded it as "none of their [the agency's] business," or decided that "they don't need to know that." Thus, respondents mobilized a set of assumptions about what questions should be asked to serve the survey's intended purpose. Respondents formed these impressions from general knowledge about the sponsoring agency (however vague and inferential), explanations given to them at the start of the survey, and prior experience with similar data collections.

Once the subject matter of the data collection went beyond the respondents' assumptions about what they should be asked, they told us they often refuse to answer. Such questions are considered "unnecessary," "nosy" or "a fishing expedition." The requested information might not be considered sensitive, but the question broke the boundaries to which the respondents believed they had agreed. This is one reason why respondents object to questionnaire supplements which are included in some panel surveys. Most of our Phase I respondents had been exposed to a supplement in CPS asking them about tobacco use. Some respondents were uncomfortable with these questions because the official topic of the interview was employment, and tobacco questions were unrelated to that end. This is also why respondents refused to answer certain questions in market research surveys, when the subject is expanded beyond a product or service they had actually used.

It is worth noting that a good deal of the recent complaint about the "intrusiveness" of the long form in Census 2000 may have had this form. The advertising campaign that accompanied the census and general discussion of the event was effective in informing our respondents about one legitimate purpose of the Census: that of counting everyone in the United States. Therefore, questions about commutation or housing (which might have caused no difficulty in a survey with different primary purpose) appeared unrelated to the count. Thus, some long form questions failed respondents' test of "relevance" and were perceived as breaking a privacy boundary.

[4] In fact, information collected by a seller in the course of a transaction legally belongs to the seller, although we never encountered a respondent who was aware of this.

4.2.4 Relevance of questions to the self

Respondents were concerned with the relevance of questions to the particular circumstances of their lives and interests. Thus, one respondent told us that she would not participate in a survey by a school board, because she does not have any children. Other respondents refused political polls because they were not interested in politics, and did not know much about it. In both of these instances, the refusals were based on the notion that their answers could not be "helpful," and would therefore be irrelevant to the purposes of the survey.

4.2.5 Assessing the consequences of giving information

Another important way that respondents assessed questions was by examining the consequences that might flow from providing certain information to certain sponsors. These can be generally described as benefits and risks. The negative consequences of giving information are described here as "risk" rather than cost, because respondents seemed more concerned with harm than with effort or expense.

4.2.5.1 Benefits Possible benefits are an important reason for respondents to provide information. These respondents were familiar with trading information for particular benefits in many venues, including insurance and job applications, applications for loans and mortgages, and paperwork for social service agencies. In these circumstances, respondents told us over and over, "you have no choice" but to give the information, even if the questions seem nosy or sensitive. It was our impression that poor people in our society are very used to trading information for benefits, but respondents in all classes are familiar with the experience.

Respondents trade off risks and benefits in a wide variety of less critical situations. One of our vignettes described supermarket "club" cards, which collect marketing information in return for discounted merchandise. The trade-off was apparent in how respondents reasoned about this situation. Even if they were highly protective of information and conscious of privacy, they did not think that the risks (primarily getting more junk mail) outweighed the benefits (store coupons).

4.2.5.2 Altruistic Benefits Benefits are not always seen as personal gain. Certain altruistic benefits of providing information are also taken into account by our respondents. People are motivated by a sense of doing a good for their community, however they define it, or for society at large. When we asked why people participated in the recent census, probably the most frequent answer was in terms of "being a good citizen" or because the census was "important" for their local area or ethnic group. Another value is described as "having one's voice heard" or one's life circumstances represented in the data. For example, members of ethnic minorities, or those who identified as single mothers or gays, sometimes said they participated in surveys in order to make sure that their group was represented in the data. These abstract concepts appeared highly salient to some respondents, and were sometimes enough to counteract potently resented risks. For

example, one respondent who was a firm believer in the evils of the federal government collecting information had not only answered the census herself, but talked to her neighbors about its importance. She made an exception for the census, because of her understanding of its importance to her town.

However, understanding these benefits in the abstract does not necessarily mean that respondents will agree to provide information. If respondents perceive the benefits as too marginal, or themselves as unable to share in them, they may refuse to participate even if the information requested is not particularly private. For example, one homeless man in Oakland told us that he hadn't filled out a census form in 2000, although he had learned that the census brings money into the community, because he was unable to find a shelter program that would accept him. Others told us that they didn't believe that money would come into their disadvantaged communities as a result of the census, regardless of what was promised, on the basis of past experience. They were extremely skeptical of the advertising campaigns that had stressed such benefits. As one respondent put it, they had cooperated with the census in 1990, and they had yet to see any schools built in their neighborhood.

4.2.5.3 Risks Negative consequences seem relatively more salient to our respondents than positive ones do. To a great extent, the possibility of negative consequences controls what data respondents feel comfortable in revealing. Four important kinds of consequences stand out in our data: physical danger, loss of control of data, fraud, and getting in trouble with government authorities.

4.2.5.4 Physical danger as a risk Some respondents mentioned the possibility of physical danger as a consequence of providing information. Work and home address are sometimes considered highly private because of the possibility of a stalker finding the respondent. We also heard that certain elderly respondents prefer not to answer in-person surveys, because they are afraid of having anyone come to their doors.

4.2.5.5 Fraud as a risk The consequence about which our respondents worried most was fraud. Information that could be used to access or defraud financial accounts was almost universally highly protected. It included social security numbers, credit card numbers, bank account numbers and the like. Collectively, this information was often termed "your numbers."

This worry focused on the actions of criminals. Some respondents had heard of cases of identity theft, in which a fraudulent individual creates debt in someone else's name. Many had also personally experienced difficulties in which financial information was misused. Anecdotes about problems with Internet purchases, credit accessed by strangers, and hucksters trying to elicit social security numbers over the telephone were not uncommon. Most people believed that eventually they would be able to "clear their names" but realized that it might be a lengthy and costly process. The potential to lose money or to have debts illegally created in one's name was perhaps the most serious consequence with which our respondents were concerned.

4.2.5.6 Loss of control of data as a risk Loss of control of data also concerned our respondents. They were concerned that data given willingly to one source may be transferred to third parties without their knowledge or consent. Upswings in junk mail and phone calls are the most frequently mentioned evidence that this has occurred. Thus, respondents note that house purchasers are flooded with advertising for gardening tools, and new parents with calls from diaper services, etc. Overstuffed mailboxes and interrupted dinners may seem relatively minor consequences, but represent real irritants to respondents, because, as they say, "now you know your data is out there."

In fact, it was the very uncertainty of having data "floating around" which could be distressing. Respondents expressed this anxiety in terms like "you don't know when it will come back to haunt you." Or another: "it's just the unknowing. You don't know what could become of it." Concerns about loss of control of data are at the root of many of our respondents' attitudes towards providing information over the Internet. Despite wide differences in technological knowledge and experience, almost all of our respondents perceive some risk in providing data across the net. Concerns about data being "out there" appear to be intensified by this technology.

4.2.5.7 Data sharing as a perceived risk. In our second research protocol, we attempted to address ideas about data sharing more specifically, by building in questions about information technology, vignettes which addressed the possible use of administrative records to replace a survey, the sale of data by commercial enterprises, and by probing for ideas about government keeping and sharing of data.

The reaction to data storage was not universally negative, even if the respondents understood that the data could be sold or shared. We have previously described a vignette about grocery store "clubs" that collect data on purchases. Even highly privacy conscious individuals did not see these databases as problematic. Respondents reasoned that: 1) they didn't care who knew what groceries they purchased; 2) it was worth trading the information for the discounts; 3) "embarrassing" purchases could be paid for in cash; and 4) nothing bad had happened yet. Thus, it appears that the kind of data involved and the manageability of risks control attitudes towards data sharing.

However, the transfer of more sensitive data, especially without permission, can elicit a very different response. In our second protocol, we included a vignette which described a pharmacy which was sharing customer prescription information with drug companies and researchers. Prescription drugs are more sensitive than groceries, and may in fact be thought of as confidential information. Transfer of these data was often rejected by respondents, (if they could see an alternative way for the vignette character to get necessary prescriptions filled). They said they would particularly resent this practice occurring without prior notification and permission. It is interesting to note that most people believed that this information belonged to them and that they had rights to control its disposal.

4.2.5.8 "The big computer" The vignettes described above elicited reactions to specific instances of data sharing. However, our discussions revealed a sense in which data sharing was thought to be at once larger scale, vaguer and more threatening. We came to refer to this as the "big computer" theory. Essentially, this is the idea that data available on a computer in one location will (sooner or later) be available in all computers everyplace. The imagery associated with this cultural representation is supported by the media. Respondents alluded to news stories involving hackers accessing data from private industry and from secure government agencies. They also mentioned movie images picturing police or federal agencies tapping into huge databases to reveal a person's history, face and whereabouts.

Respondents' awareness of the actual public accessability of data varied somewhat. Some respondents were quite knowledgeable about actual public sources of information. They pointed out that in most of the states where we interviewed, social security number is used as a driver's licence number, and is copied down everywhere. Others mentioned that names and addresses are available on voter registration lists and that it is easy to get credit information from credit bureaus. They knew or had heard that it was possible to look individuals up on the Internet. Thus, to some extent, respondents' belief in the public accessibility of data about them is based in fact.

Although "big computer" ideas are not exclusive to views of government, they are most commonly believed in that venue. The exchange of data between agencies is often seen as happening easily and quickly through centralized files. (One respondent described for us the computers sitting on our desks in the Census Bureau where we could pull up data about anyone in whom we were interested). In general, our respondents believe "government" shares data among different agencies and levels of government, regardless of any promises of confidentiality that are given.

They assume that information they give will make its way back to interested authorities. Thus, it is assumed that the INS will find out about illegal aliens and the IRS about tax evaders, regardless of who is collecting the data or what promises they give. One of our vignettes described an undocumented immigrant who is asked about his immigration status in a survey and is promised confidentiality. Most of our respondents thought that answering truthfully would be very risky. (Two of them thought he should answer the questions truthfully because they thought he should be deported, and this information would facilitate it). A similar vignette described a man who fixes cars off the books, and is asked about his income in a survey. Again, most respondents did not think that he should reveal his cash payments in any survey, because they would then be available to the IRS.

In particular, law enforcement and the courts are assumed to be able to get whatever data they want. Respondents sometimes recounted anecdotes of local police being able to find individuals by using data they had given to other agencies (such as housing, motor vehicles, or social service agencies,) and this is taken as proof of widespread information sharing. While this is seen as a major risk factor by some respondents, others think it may be proper, because it allows malefactors to be caught, (or in one case, a respondent's runaway son to be located). Belief in police

powers to get information may not entirely invalidate a respondent's belief in claims of confidentiality. Assurances of confidentiality may be assumed by respondents to apply to normal circumstances and to the behavior of average agency personnel. However, if a high level employee of a law enforcement agency demands data normally considered confidential, such respondents have no doubt it will be made available.

The belief that all government information is available to all government agencies had some interesting ramifications. Although many of our respondents try to be careful with their social security numbers, they often thought there was no additional risk in supplying it on a government survey. As they pointed out, "they *[the government]* gave it to me in the first place." Many respondents were puzzled by why a government survey would ask for social security number in the first place. One respondent thought that asking for social security number might make him suspicious. If they have to ask social security number, he reasoned, they may not actually be a legitimate government agency.

Government data sharing is implicit in ideas about government "tracking" of individuals, alluded to previously. However, the purposes of this tracking are really not clear to our respondents. When probed about the specific purposes of this government tracking, two themes emerged. One was an interest in keeping track of the location of individuals (perhaps rendering them more accessible to the police). The other was in monitoring wages, assets and other financial transactions, presumably with the purpose of uncovering lies on income tax forms.

4.2.5.9 Administrative records use Concerns about data sharing have a strong effect on respondents attitudes to the idea of administrative records use. One vignette described a government survey in which a character could either fill out 40 questions on each family member or give permission for the same data to be acquired from other agencies. The reaction to this vignette was interesting. Forty questions seemed burdensome to most respondents. But the idea of allowing an agency carte blanche in one's data files was disturbing, and many respondents who were initially tempted by the saving of time changed their minds. Others were angered by the suggestion, and they were inclined to refuse both the survey and the permission to look at other records.

The risks they associated with the administrative records use were the following:

- Reservations about the accuracy of data already in the files. They are afraid of "mistakes" in the data, or out-dated information about them being perpetuated. Some worried about being held responsible for the incorrect data.

- Discovery of contradictions between various sources. Respondents frequently modify the data they give for specific purposes, so they are aware that they may not have reported exactly the same information in every venue. These respondents regarded the suggestion to use data from other records as an opportunity to "check" their answers.

- Some respondents think it is more risky to have all data in one central location, and deliberately follow a strategy of telling only part of the data in any one venue.

Because of these concerns, most respondents felt that they had more control and less risk by refusing permission to get data from administrative records sources and filling out even a burdensome questionnaire themselves. A few respondents, however, felt that the choice as presented was unrealistic. They already believed in the wide availability of data between government agencies, thus they believed the administrative records data would be checked even if they refused permission.

Respondents are not just afraid of intentional data sharing by agencies that collect it. They are also concerned about data being "shared" because of the bad behavior of individuals. Hackers are a prime example of this concern. Respondents believed that even if an agency has an official policy of confidentiality, people with bad intentions can access these computers from outside and steal data that everyone thinks are protected. This is one reason that assurances of confidentiality were not completely convincing to respondents, since they did not believe that reputable organizations are effective at protecting themselves from these outsiders.

4.3. Managing information

Respondents' concerns about information led them to attempt to control what others can find out about them. Since respondents are highly concerned with fraud, most salient are their efforts to protect "their numbers." When we asked if there was anything that people did or avoided doing to protect information about themselves, respondents told us about cutting horizontally through the numbers on out of date credit cards, carefully destroying "preapproved" credit card applications, using shredders for bank or credit card statements, whispering their driver's licence number (hence, social security number) in stores, hiding check deposit slips, and always hanging up on anyone who seemed to want this information over the telephone.

When the requested information in a survey or application was beyond respondents' comfort level, some of them recommended asking the sponsor "if you really need that information," or if you "have to give those answers." This suggestion often arose if the questions provided access to a benefit: respondents explained that they were trying to establish whether they could leave the answer blank, and still be eligible. For example, if prior medical conditions might affect access to insurance, these respondents will try to negotiate how much they are required to reveal. But other respondents differed: they believed in the wide accessibility of the information, and reasoned that the authority would inevitably uncover any omitted information, and they would be subject to additional penalties.

Respondents also use "don't know" and "not applicable" options to manage information, especially if they think that the questions are irrelevant to the legitimate purpose of the questionnaire. Another commonly mentioned technique for protecting information was lying. We were struck in these interviews about how often respondents reported that they lied in response to requests for information. This is particularly true if the questions being asked are beyond the boundaries of what the respondent considers to be legitimate. Respondents appeared to feel little or no

compunction about telling these lies. Sometimes these lies presented a modification of the truth, such as reporting only part of off-the-books income. In other cases, respondents presented outright untruths, such as giving an incorrect zip code or phone numbers to prevent marketers from finding them. One respondent enjoyed making up bizarre answers to mall surveys, as a kind of game.

4.4 Diversity in Privacy Beliefs and Behaviors

4.4.1 Technological Awareness

Because of the importance of computers and data sharing to respondents' anxieties about information, we wanted to interview persons with a variety of levels of technological sophistication. We had many respondents who had never used computers or the Internet, (although most had friends and relations who were users). We also recruited some respondents that worked in the computer industry or in data mining. On the whole, the ideas that govern the decision to reveal information did not seem to differ too greatly between the two ends of the spectrum. That is, there is considerable commonality about assessing sponsorship, relevance, and attending to tradeoffs between risk and benefit. However, some differences did emerge.

The more technologically sophisticated respondents had a somewhat different attitude towards the Internet. They were more familiar with it, and one respondent told us that she felt comfortable answering questions on the Internet since she "thinks better on a computer." However, the technologically knowledgeable respondents, like other respondents, believed that information supplied over the Internet was at risk. For example, one respondent employed as an analyst in a commercial data mining firm refused to make purchases over the Internet and did not want to give sensitive survey information in that mode either:

> "Maybe my answer would be different five years from now, but the security is still not very good. When I type messages, when I type email, I assume that everybody in the world is reading what I'm writing...so if I transmit something I just assume everybody's seeing it."

However, other such respondents were more certain of their ability to determine a secure site, and were therefore more willing to use the Internet for important transactions. They told us they looked for a privacy policy on a site. (Although they might not read it completely, they liked that it was there). They looked for specific icons indicating secure sites (a key) although this was not entirely reassuring, since icons are easily made. A few mentioned checking the encryption programs which were in use before supplying information. But even these assurances did not convince them that the Internet was entirely without risk, and they too were primarily worried about fraud.

Because this group of respondents did not look at the Internet as entirely secure, their attention focused on the ways in which it is possible to mitigate negative consequences if they occurred.

The strategies they had developed for this were primarily what distinguished the more technologically sophisticated respondents in our data. First, they were very aware of the policies protecting consumers over the Internet, which led them to choose certain credit cards or retailers which eliminate all financial liability related to fraud in web transactions. This meant that, like less sophisticated consumers, they preferred dealing with organizations which had good prior reputations. Second, they appeared much more willing than our other respondents to divest themselves of compromised identifiers. This strategy was called "ditching" by one respondent. These more sophisticated respondents consider it relatively easy to "ditch" information such as a credit card number, a bank account, an email address, or a post office box. The process of getting these new identifiers did not seem like a "hassle" to those who embraced this strategy.

Another difference that is worth noting is that these respondents had a greater awareness of the existence of extensive private, as opposed to government, data bases on individuals. Unlike our other respondents, they mentioned information maintained by large corporations on their customers, industry wide data bases such as a "pooled cooperative database of catalogue merchandisers" and a shared industry wide data base with information about 60 to 70 million households. The extent of this private data sharing sometimes made respondents who were aware of it give up on attempting to control information at all. It is interesting to note that this respondent attributed his attitude to his job working with a large commercial data base.

> "Yeah, well, so what? I mean, what are they going to do about it?...I would imagine they must rent that information out or something... because they must, they just must. Everybody's doing it to everybody else, they must...Well, here is my personal belief...Everybody, anybody who has any information about me will sell it if they can find somebody to buy it. So asking for consent is kind of meaningless...I mean the sense in which they're disclosing that they're doing it, I mean that element of this seems unnecessary to me because I'm assuming everybody is doing it. And, knock yourself out, is my attitude...I mean before I worked here I never gave it an ounce of thought. And I certainly wasn't aware of the extent to which people are selling information to each other."

4.4.2 Group differences

One original aim of the research was to discover differences in cultural beliefs and attitudes about privacy which might affect the response patterns of particular groups. We had expected to discover differences in the definition of privacy or different patterns of protected information among these groups. For this reason, recruiting stressed Native American, African American, Asian, and Latino respondents. On the whole, however, the emerging picture does not indicate significant differences in approach to privacy that can be ascribed to ethnic cultures. That is to say, faced with the same data collections and similar risks and benefits, there is considerable commonality in the way in which these groups approach revealing information. Thus, respondents in all groups process decisions about revealing information by thinking about sponsorship, relevance, risks and benefits. Fraud and loss of control over data are important concerns. It

should be noted that our research protocols tended to stress decision making in practical situations where data were being collected by a commercial or governmental organization. This focus was appropriate to the research, but was not designed to elicit differences in the social construction of privacy in interpersonal interaction within the home or in community contexts. A complete account of ethnic differences in the interpersonal aspects of privacy awaits further research.

Nevertheless, we are able to suggest some patterns in which these groups may be said to diverge from the common account already rendered. These include:

- the effects of the experiences of particular groups with government on their attitudes to data collection;
- the contrast between privacy in communally based cultures and individually based cultures;
- the effects of social class on decisions about privacy; and
- different privacy sensitivities.

The descriptions below should be considered as suggested hypotheses for further research.

4.4.2.1 Experiences with government The historical relations between groups and the government appear to have had a strong effect on attitudes towards data collections. For example, among Native Americans, beliefs in government tracking of individuals seemed to be relatively common. One view was that "keeping track" or "keeping tabs" on the population was the purpose of the decennial census. This seemed to be redundant effort to one Native Americans respondent, since the federal government "knows where I am, they can find me." One respondent alluded to this by mentioning the Certificate of Degree of Indian Blood, given by the Federal Government to enrolled members of federally registered tribes, which controls access to such benefits as the Indian Health Service. Another respondent mentioned a belief that the FBI maintains files on Indians, especially if they had connections with certain Indian political movements.

The "tracking" idea seemed particularly troubling to a few Native American respondents. Here is an example of a respondent who found the census intimidating:

> "I really feel intimidated that I have to let the government know where I live, what I do, how many in my family. Almost, basically, running my life now. And I don't like it. I hate it. Somehow I think that with all the computer technology they have now, that they could track you...that's actually how I feel, like in the wild how they tag the animals and then they could tell where you're at, how far you've traveled in a given time, and ...with the census thing, I don't know why they do that. I still don't know. Why they have to track every body like that, why do they have to know who lives in your house, what do you guys do every day..."

It should be noted that these negative views of the census were more common among one group of Native American respondents than in the other we interviewed. The group where these ideas were more salient lived in the Oakland area, and its members were somewhat older than the group interviewed in Los Angeles. Further, they were part of a group which was formed initially in the 50's and 60's as a result of a federal relocation project. Some of the respondents were children of Native Americans who had been removed from their homes, and sent to school in the Bay area. This experience is still remembered and resented. By contrast, the group in Los Angeles was younger, and generally better educated. The attitudes of the second group do not show as great a sensitivity to the relationship between the federal government and Native Americans. Most were quite positive about the Census.

Other attitudes towards the government also reflect particular relationships between the government and a local ethnic community. In particular, problems with the INS were mentioned by Latino and Asian respondents as frequently heard concerns about the Census. Some of the Latino interviews were done in San Diego, and respondents there reported negative interactions with agents of the federal government because of the proximity of the border. A respondent had been detained crossing back into the United States after a visit to Mexico while officials demanded documentation including birth certificates, pay stubs and rent receipts. This makes respondents feel powerless. As she remarked, "they're federal agents, they can do it."

Perceptions of our government may be mitigated by contrast to more repressive governments in countries from which our respondents emigrated. For example, one Cuban who had recently come to Florida found this country more private than her homeland:

> "In this country I think there's quite a bit of privacy in people's lives. That's one of the things I like best. Because in the country I come from, they look even into the toenail of your big toe. So you want to get away from there because there's so much they want to know about your life. They have you under surveillance. You feel completely asphyxiated, like you have no privacy even in your own home."

4.4.2.2 Communally based cultures. There is some suggestion in these data that concepts of privacy may be influenced by the sense of community which exists within a particular group.
This was particularly striking in terms of a group of respondents who were immigrants from India. Most respondents in this research believe that privacy has decreased in recent years, but these Indian respondents sense more privacy in America than in India. There, family and acquaintances expect to know all sorts of personal details about ones' life, such as income or plans to have children, and apparently have wide rights to inquire about such matters. For example:

> "In the U.S .everything is pretty private. There is more privacy in this country than back home in India. There is a lot of socializing that goes on in our lives back home. So, information passes around pretty fast. And it's quite common...like for example, the kind of money you make or the sources of our income or the relations we have with other people, general things. These things are open to a certain extent in the community."

In general, they said that they liked the new sense of privacy. They also noted, however, that protecting information was more of a concern in the United States, and that they had learned to worry about it after coming here:

> "I never paid so much attention to privacy before coming to America, and I never actually thought of privacy as such a show-stopping, life-critical thing. But after watching so many Hollywood films, I probably think that privacy could be protected closely...It's a big deal, is what I'm feeling right now..."

In dealing with decisions to reveal information, these respondents appeared to be primarily influenced by their new concerns, and in fact sounded much like the rest of our respondents in describing the decision to divulge information.

A different sense of communally held information also emerges from our interviews with a group of poor African Americans. Their assessments of whether particular information is in the public or private realm tend to take account of the ease with which information spreads within tightly knit communities. Thus, information which might be considered damaging, such as having more than the allowed number of residents in an apartment, was not classified as private, because everyone in the neighborhood knew about it. It should be noted that just because this information was considered in some sense public, it did not mean that they were willing to reveal it in a government survey.

4.4.2.3 Social class Social class appears to have a considerable impact on the responses to matters of privacy. First, as we have already mentioned, the poorest respondents may not be influenced by messages couched in terms of the benefits, because they may have ample evidence that resources tend not to be funneled to them or to their communities. Second, being poor may restrict one's options in attempting to protect privacy. This became clear in the reactions of most of the group of poor African Americans to a vignette in which the central character has to decide what to do about a pharmacy which is selling information about drug purchases. The typical response of the more affluent respondents was to suggest changing pharmacies. However, these poorer respondents sensed that this was not possible if no other pharmacy existed or gave credit in the neighborhood. In a sense, this is an example of trading information for benefits, to which we have already referred. It should be kept in mind that when options are restricted, protecting privacy may take a lower priority to other matters.

Another difference that social class may create in these data occurs in reactions to various modes of question administration. In general, more affluent respondents preferred modes of administration which allow them to stay in control of their time and living space: thus, mail (and for some, the Internet) are preferred modes. However, among our poorest respondents, face to face interviews tend to be more highly valued. Respondents say that they like to be able to assess an interviewer in person, in order to be able to decide if they are trustworthy. They have, perhaps, more confidence in their ability to read individuals than to determine if written promises of

confidentiality are dependable. In addition, some respondents see the interviewer as a source of explanations of difficult material and a possible helper if giving the information proves somehow damaging.

4.4.2.4 Different privacy sensitivities Matters of sexuality may be a salient privacy concern with some Latino respondents. There is evidence that the Spanish term "privado" (private) tends to elicit associations with sexuality which the English term "private" does not. Thus, "privado" was defined as "Things about a married couple that no one should intrude into," "Intimate things. In couples" and "It is something you shouldn't do. Like something forbidden." This should be taken into account in Spanish translations of privacy statements, etc.

Our Native American respondents indicated certain sensitivities which were not mentioned by other respondents. Being asked questions about their children, particularly their names, was mentioned by some as problematic. This may have been due to the history previously alluded to, of children being separated from their families by the government. In addition, matters of religion and spirituality were mentioned by these respondents as issues they wanted to keep very private. They expressed a concern that the wider society might find their religious practices odd or different, and were thus very concerned with keeping them within the family or the community. The sensitivity to revealing names may also have religious connotations for some Native American groups, where names may have a sacred connotation.

5. CONCLUSIONS

5.1 Diversity and commonality in privacy beliefs

Overall this research indicates that there is a wide common area of agreement between respondents in the way that they make decisions about requests for information. Some of our major findings, such as the concern with sponsorship, relevance and risks and benefits, can be found in every group we interviewed. All are concerned with the possibility of fraud, and give high priority to the protection of financial resources. Suspicion of the security of the Internet occurs in all groups, regardless of the degree of experience with the mode or technical expertise with computers. Thus, diversity does not stand as clearly as the commonality as a result of this research.

However, thinking about the diversity we did find is useful. First, we need to look for attitudinal and behavioral variation within the very broadly defined ethnic categories which tend to structure our research. From this point of view, analyzing the beliefs of "Native Americans" or "Asians" is less revealing than defining the research unit as, for example, "urban Indians in the Bay area" or "middle class immigrants from India." A complete cultural account of reactions to privacy in surveys, structured in this way, would require a much larger research project than we were able to do here.

Second, the differences we see between groups may reflect factors such as social class or contingent aspects of a specific historical relationship with government rather than ethnically specific beliefs and definitions. Thus, technically sophisticated and middle class respondents resembled each other strongly, despite their membership in different ethnic communities. Our results also suggest that there may be more similarities between the disempowered of various groups than between the members of various ethnic groups. Future research is necessary which examines these dimensions of difference directly, rather than assuming differences to be coextensive with ethnic community. This strategy highlights similarity of social situation and relationship to power as explanations for privacy behaviors, rather than emphasizing ethnic differences from a "mainstream culture."

5.2 The cultural understanding of privacy

The control of information is central to our respondents' understanding of privacy. This stress on control is consistent with general American values. Loss of control over information is resented or provokes anxiety. Respondent's sensitivities to control of information determine their attitudes towards issues such as data sharing between agencies, and affect attitudes towards different modes of questionnaire administration.

Because privacy is evaluated situationally, it is not possible to define in advance a set of topics which are perceived as breaches to privacy in all situations. As we have indicated here, perception of a legitimate need for the information is a critical factor in situationally determining whether the request for information will trigger privacy concerns. Respondents require a sense of the legitimate, beneficial uses of information before they release it. Benefits to a particular community or to society as a whole do serve as motivations, and as such are critical to communicate, (although, as we have seen, this can backfire with the most socially marginal individuals). This leads to the conclusion that explaining the uses of the data to respondents merits considerable attention on the part of questionnaire designers. We suggest that these explanations must be on the level of the specific information which is requested and not on the level of the entire data collection. Further, the inclusion of topical material which is only distantly related to the publically known purpose of the data collection may be a risky strategy.

Respondents also assess risks to themselves. It is important to note that assurances of confidentiality are generally not enough to counter a defined risk, especially when the stakes are considered to be high. The public perception is that data are widely shared between government agencies.

Explanations of confidentiality should be changed to reflect this concern. These explanations should include more than descriptions of official policy, but should also explain the way in which data are protected from intrusion by outsiders. However, we found little that would serve to counteract this belief for people who had a great deal to lose.

These data suggest another limitation of assurances of confidentiality: if the privacy reaction results from a failure to see the legitimate purpose of the data collection, assurances of confidentiality are actually irrelevant to encouraging participation. (If a question is "none of our business" it really doesn't help to assure the respondent that we won't tell anybody else.) Thus, reactions to privacy cannot always be managed with assurances of confidentiality. Other avenues, primarily better explaining the legitimate need for the information, should be pursued.

6. RECOMMENDATIONS

- **Because privacy judgements are situational, it is not possible to create a list of items that will always or never be considered private.**

- **Because the sense of intrusiveness of questions is situational, be careful how disparate topics are combined in one survey setting (such as topical modules or supplements.)**

- **Include the idea of having one's "voice" heard in motivational material for minority groups.**

- **Through media coverage of other agencies and organizations, respondents are aware that fraud may occur through the action of individual employees. Describe the Census Bureau's internal controls on the handling of data in explanations of confidentiality.**

- **Because respondents' comfort with questions rests on their assessment of the sponsor's legitimate right to know the information requested, provide good, understandable explanations of why these data are needed and how they will be used.**

- **Further research is necessary to assess changes in privacy beliefs and behaviors resulting from recent events.**

References

Gerber, E. *1994. Hidden Assumptions: The Use of Vignettes in Cognitive Interviewing."* Paper presented to the American Association for Public Opinion Research, May *1994.* Joint Proceedings of the American Statistical Association, *1994.*

Appendix I: Privacy Language

Expressing comfort
Open, nothing to hide,
trust, trusting, confident, secure, protecting, controlling access

Expressing privacy concerns
wary, leery, skeptical, unsure,
paranoid, paranoid factor, makes you nervous, hesitant, intimidated,
delve, prying probing, in depth, pushy, gone too far, digging out [information]
intrusive, *intrusa,* getting into your business, going behind your back, deceivingly

Expressing boundaries:

Personal boundaries: touchy, sensitive, private, *privado,* firewall, brick wall around the heart, confidential, quiet, to herself, not open to everybody, keep things to yourself, inner circle personal, personal space, personal dealing personal finances
control, ownership of yourself, boundaries
(Negative) nosy, big mouth, pot stirrer, loose-lipped, *chisme, chismosas,* meddle
leave it alone, I can't tell you or I'll have to kill you, what's it to you? don't go there, none of your business, you don't need to know, leave me alone

Family boundaries: keep it in the household, this stays in the house, keep that behind closed doors, just between us, airing your dirty laundry, putting it on the street, that's one of our family matters, family issue, keep it in the low down

Relating to data sharing:
central information bank, government files, big computer file, master list, track, tracking
established company, reputation, fly by night, secure site, secure server, access, distributed, dispersed,
divulge, open the door to your information

Appendix II: Semi-Structured Interview Protocol

Introduction.

Hi, thanks for agreeing to help. I'm talking to people to discover some of the reasons why people participate in the census, and some of the reasons why they don't. I'm interested in talking with you informally about your opinions and feelings about the census and answering questions. This interview should take about an hour and a half. As we discussed previously, anything we discuss today will be strictly confidential, and your participation in this interview is completely voluntary.

I. Census debrief
Let's start by talking about your experiences with the census.

1. Did you receive your census form? Did you receive the long form or the short form?

Had you heard about the census before it arrived? -- what had you heard (advertisements)?

Did someone in your household fill out the form and return it? -- who/ or did you talk to someone from the census (enumerator)?

Did you have any problems or concerns about the questions or filling it out?

Alternate probes: Did you hear anybody else talk about problems or concerns? Did you hear anybody talk about not liking the questions?

2. What do you think might influence people to participate in the census or not participate? 3. If

people don't respond, why do you think they don't?

II. Experiences with Other Data Collection.
I'd also like to ask you about other requests for personal information. We're not just interested in surveys you might have participated in recently, but all types of questions asking you to provide details about your everyday life.

4. Besides the census, have you been asked any questions over the phone lately, or did you fill out any forms, or did any interviewers come to the door? What were they?

5. How do you decide if you will give information in these instances?

III. Mode Preferences Section

6. If you could choose the way to give your answers, which would you prefer: a face-to-face interview, a phone interview, a mail-back paper interview, an internet interview? Which would you like the most? Why?

Rank order responses from Rs

7. If the survey had a interviewer come to your door in person, would you would prefer that the interviewer be someone who lives in your neighborhood or someone who lives outside your neighborhood? Why?

IV. Privacy: Issue Awareness, Self-Assessment

8. Some people care a lot about keeping their life private, while others do not care as much. What about you? How so? Can you tell me more about that? (Expansion probes)

Possible probe: Would you say that your spouse/partner/relatives are more or less interested in privacy than you are?

9. Do you think there is more or less privacy in people's lives today than there used to be, or is it about the same?

What factors have increased/decreased privacy?

10. Do you think that technology has made a difference in peoples' privacy? What technologies? If necessary: What about computers, have they made a difference in peoples' privacy?

V. Vignettes *VARY* THE ORDER as you present these.

Now we'd like to turn to some different kinds of tasks. First, I have some situations on these cards. Each one describes a decision that someone has to make, and asks what that person should do in that instance. The situations are ambiguous, and there is no right or wrong answer. We're interested in YOUR opinion of what that person should do.

1. George wants to buy a gift. He finds an internet website which has what he is looking for. It asks him to enter his credit card number to complete the sale. Should George provide this information? Why or why not? (Do you purchase items from the internet?)

Do you or someone in your family use a computer? Do you use the internet?

2. In order to get a good discount on a car on the internet, Janie must fill out a computer form providing the names and email addresses of four family members or friends. Should Janie provide this information? Why or why not?

3. Tamara receives a survey in the mail from the federal government that asks for her SSN. The survey promises her confidentiality. Should Tamara provide this information? Why or why not?

4. Ivan is an undocumented immigrant. A government survey asks him about his residency status in this country. The interviewer promises him complete confidentiality. She explains that personal information is never released, and that his name will not be connected with his answers in any way. Should Ivan provide this information? Why or why not?

5. Paulette has gone to the same pharmacy for a long time and likes her pharmacist. She discovers that the pharmacy keeps track of information about customers' drug purchases, and sells it to other businesses and research firms. Should Paulette continue to go to the same pharmacy? Why or why not?

6. To supplement his income, Andy fixes cars for friends and neighbors in his backyard, and asks that they always pay cash. A survey interviewer asks Andy for his earnings from all sources of income. Should Andy answer these questions? Why or why not?

7. Sandy receives a letter from a government agency. The letter says she can either fill out a questionnaire with 40 questions on each person in her household, or with her permission, they can get the same information from other agencies. Should she give permission or fill out the questionnaire? Why or why not?

Probe: Why do you prefer that choice?
Probe: How long was the longest survey you've completed?
Probe: What would you say if the mail-back survey said "required by law"?

8. In order to get their child into a good pre-school program, Sue and Bob have to answer a lot of questions about their family life, such as how often they quarrel and how they discipline their children. Should they give this information? Why or why not?

Possible Additional Probe: What if it was a local charity asking those questions?

VI. Privacy Schema
Now I'd like to discuss some of the ways in which people express their ideas about privacy. *(Choose no more than 5 terms/phrases that the respondent has used, and discuss them with the respondent.)*

When we were discussing the situations earlier, I noticed you said xxx. Could you give me some examples of xxx? It doesn't have to be something that really happened, just a general example.

Only ask for abstract meaning probes if the respondent seems able to answer them.

Here are some other expressions people use to talk about privacy. Can you give me some examples to explain these? (Only present terms that the respondent has not discussed so far.)

none of your business
private
keep it in the family
personal
intrusion of privacy

VII. Confidentiality Statements: Card Prompts. Questions below.
VARY THE ORDER as you present these to interviewees.
Now I'd like to show you some statements about confidentiality that may be used in some government surveys. Please read each one and tell me what you think of it. Tell me if there are phrases you like or don't like, or phrases that raise questions.

"Your privacy is protected by law (Title 13 of the United States Code), which also requires that you answer the questions. That law ensures that your information is only used for statistical purposes and that no unauthorized person can see your form or find out what you tell us, no other government agency, no court of law, NO ONE."

What is this trying to tell you?

Do you think this is easy to understand?

"Only persons sworn to protect the confidentiality of your information can see your form. No one else will be able to connect your answers with your name and address. No one, not even Census Bureau staff, are permitted to use your

information other than to conduct this survey. In the course of this survey, we may combine your answers with information that you have given other agencies to enhance the statistical uses of the survey data. This information will be given the same protection as your survey information."

What is this trying to tell you?

What is meant by 'combine your answers with ....'?

Do you think this is easy to understand?

VIII. Confidentiality phrases:
I have some of the language used in these statements, and similar statements, on these cards. I'd like to discuss each phrase.

What meaning do you get from it when they say:

"Strictly confidential"
"Confidential by law"
"Statistical purposes"
"only summary data will be published and made available..."
"we never release information that would disclose your identity."
"no one else will be able to connect your answers with your name and address"

IX. Demographic Information
Finally, I want to ask you five short demographic questions

What is your: Sex Age Ethnicity/Race _____

Occupation/Work _____ Education _____